High-Throughput Cloning for Proteomics Research

Sharon A. Doyle

Running title: High-throughput Cloning

Keywords: Ligation-independent cloning, T7 expression system, pET vector, hexahistidine tag

Sharon A. Doyle, Ph.D.
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925-296-5694
Fax: 925-296-5850
Email: sadoyle@lbl.gov

1. Introduction

Many proteomics initiatives require the production of large collections of expression clones. While traditional methods of cloning, such as restriction enzyme-based cloning, have been well established on a small scale, they are not adaptable for a high-throughput environment. Methods that facilitate cloning in a high-throughput manner are vital to the success of these initiatives.

This chapter describes ligation-independent cloning (LIC) ***(1, 2)***, an ideal cloning strategy for high-throughput proteomics. In this system, linear plasmid vector and insert DNA are treated to generate complementary single-stranded overhangs that anneal during a short incubation. In addition to the ease of LIC due to the speed and efficiency of the annealing reaction, other attributes make it an optimal system for cloning large numbers of cDNAs. Since restriction enzymes are not used, insert sequences do not need to be screened for internal sites or modified prior to cloning. Any vector can be made LIC compatible, allowing for the production of proteins with different affinity tags, often using the same insert PCR product. The additional amino acids added to the encoded protein by the LIC sequences are minimal and can be modified if necessary, which is often important when N- or C-terminal extensions affect protein function or stability.

2. Materials

1. pET30a (Novagen, Madison, WI) or other expression vector system

2. *Eschericia coli* (*E. coli*) strain NovaBlue competent cells (Novagen, Madison, WI)

3. Oligonucleotide primers

4. BseR I restriction enzyme

5. dNTPs, HotstarTaq polymerase (Qiagen, Valencia, CA), Platinum Pfx polymerase (Invitrogen, Carlsbad, CA)

6. Mung bean nuclease, T4DNA polymerase (Novagen, Madison, WI)

7. Agarose gel electrophoresis equipment

8. LB media and LB agar plates, SOC media

9. Kanamycin

10. SDS (sodium dodecyl sulfate)

11. DTT (dithiothreitol)

12. PCR cleanup and agarose gel extraction systems (Qiagen, Valencia, CA)

3. Methods

First, a method to generate a LIC vector and prepare it for cloning will be described to illustrate how any vector system used for protein expression can be modified for LIC. Next, the generation of suitable insert sequences will be described, followed by the LIC reaction, and confirmation of positive clones.

3.1 Expression Plasmid

3.1.1 Construction of a Ligation-Independent Cloning (LIC) Compatible Plasmid

Ligation-independent cloning can be used with any plasmid vector following an alteration of the sequence using standard molecular biology techniques. The LIC region inserted into the plasmid must contain the following: a unique restriction enzyme cleavage site used to linearize the plasmid DNA (*see* **Note 1**) flanked by approximately 12-15 bp on each side that act as the annealing sites for the LIC reaction. These flanking regions are used to generate long single-stranded overhangs used in the LIC reaction by treatment with a DNA polymerase with 3' to 5' exonuclease activity in the presence of one of the dNTPs; the polymerase chews back the 3' ends of the DNA until it encounters the dNTP present in the reaction. Therefore, the sequence composition of the region flanking the restriction cleavage site must be devoid of one dNTP to allow the enzyme to chew back a minimum of 12 base pairs (***3***). Several LIC sequences have been used that provide different affinity tags and protease cleavage sites ***(4, 5)***. The efficiency of the LIC reaction should be carefully assessed when new sequences are designed, however, since lower cloning efficiencies are problematic in a high-throughput environment.

Plasmids in the pET family of expression vectors (***5***), used widely for protein production, utilize the T7 promoter in conjunction with T7 RNA polymerase to drive high-level expression of heterologous proteins in *E. coli* ***(6-8)***. To generate a LIC compatible vector that incorporates a N-terminal hexahistidine tag into the protein product and minimizes extraneous amino acids, pET30a was modified to produce pNHis (**Fig. 1**). The sequence

following the hexahistidine tag sequence was removed and replaced with 5' TCCGGTATTGAGGGTCGCTCTAACTCTCCTCTG 3' using standard molecular biology techniques. This sequence contains a BseR I recognition sequence and cleavage site, which is absent from the pET30a sequence, and the flanking sequences that are used to create the overhanging ends for the annealing reaction ***(5)***. This sequence encodes the Factor Xa site (IEGR) that may be used to cleave the hexahistidine tag and LIC sequence encoding amino acids from the protein product. A diagram showing the details of the LIC reaction (outlined below) using the pNHis vector is shown (**Fig. 2**). Several hundred micrograms of plasmid DNA were generated following confirmation of the sequence.

3.1.2 Preparation of Linear Plasmid DNA for LIC

1. Digest vector DNA (10 µg) with BseR I for 6 h at 37 °C in a 200 µl reaction volume (*see* **Note 2**).
2. Add 20 U mung bean nuclease and incubate at 30 °C for 3 h.
3. Add SDS to a final concentration of 0.01% (w/v) to inactivate the nuclease.
4. Isolate the linear vector from uncut and supercoiled vector on a 2% agarose gel at 100V for approximately 4 h, or until well separated. Use undigested vector as a negative control (*see* **Note 3**).
5. Extract the linear vector DNA from the agarose gel using standard molecular biology techniques (QIAQuick gel extraction kit (Qiagen) or similar). Resuspend the DNA in H_20, and keep the DNA concentration as high as possible.

6. Determine the absorbance at 260 nm, calculate the concentration, and store in aliquots at –20 °C.

7. Test the vector DNA to ensure that all undigested DNA has been removed by performing the LIC reaction (see Section 3.3) with LIC-treated linear DNA and no insert DNA. Following transformation into NovaBlue cells, there should be very few transformants (fewer than 10 colonies is optimal).

3.2 Preparation of cDNA Inserts for LIC

1. Amplify genes of interest using a high fidelity polymerase (such as Platinum Pfx polymerase) with primers that contain 5' extensions for LIC. For the pNHis vector, the forward primer uses the extension 5' GGTATTGAGGGTCGC, followed by the cDNA sequence starting with the second codon. The reverse primer uses the extension 5' AGAGGAGAGTTAGAGCCTTA. Note that a stop codon is added to this sequence so that the product does not contain additional amino acids encoded by the LIC sequence (*see* **Note 4**).

2. Analyze the PCR products by agarose gel electrophoresis. If products are highly pure, concentrate the sample using a 96 well PCR cleanup protocol such as QIAquick (Qiagen) (*see* **Note 5**). Alternatively, products may be isolated by agarose gel electrophoresis and extracted and concentrated using an extraction protocol such as MinElute (Qiagen). Highly concentrated PCR products produce optimal results (0.1 pmol/μl) so resuspension volumes should be chosen accordingly (*see* **Note 6**).

3.3 Ligation-Independent Cloning

1. Separate reactions are performed for plasmid vector and insert DNA as follows. In one tube, 0.1 pmol linear vector DNA (400 ng for pNHis), 2 µl 10X T4 DNA polymerase buffer, 2 µl 25 mM dCTP (*see* **Note 7**), 1µl 100 mM DTT, and 0.4 µl LIC qualified T4 DNA polymerase (1.25 U) are added to a final volume of 20 µl. In another tube (or reaction plate) 0.2 pmol linear insert DNA (*see* **Note 8**), 2 µl 10X T4 DNA polymerase buffer, 2 µl 25 mM dGTP (*see* **Note 7**), 1µl 100 mM DTT, and 0.4 µl LIC qualified T4 DNA polymerase (1.25 U) are added to a final volume of 20 µl. Incubate the reactions at room temperature (22 °C) for 40 min.

2. To stop the separate LIC reactions, heat inactivate at 75 °C for 20 min.

3. To anneal the vector and insert, combine 2 µl (0.01 pmol) of LIC vector and 2 µl (approximately 0.02 pmol) of LIC insert in a tube and incubate at room temperature for 10 min. A negative control sample of 2 µl vector and 2 µl H_20 should be included.

4. To stop the annealing reactions, add 1.3 µl of 25 mM EDTA and incubate at room temperature for 10 min. Reactions may be stored at –20 °C prior to transformation if desired.

5. Transform 1.3 µl of the annealing reaction (diluted 1:5 with H_20) into NovaBlue competent cells and plate the entire reaction onto LB agar plates containing 50 µg/ml kanamycin. Incubate overnight at 37 °C.

3.4 Identification of Positive Clones

It is beneficial to identify clones that contain the correct insert prior to plasmid purification and sequencing. The following protocol for colony PCR utilizes primers that anneal to the vector sequence, which is advantageous because one optimized set of PCR conditions is used. This allows for easy preparation, and avoids problems related to different primer conditions. In addition, a negative control of vector DNA can serve as a built-in control for the reaction. Primers should be chosen that produce a product of approximately 100 bp when vector DNA without insert is amplified. To amplify the pNHis vector, the T7 promoter (5' TAATACGACTCACTATAGGG 3') and T7 terminator (5' GCTAGTTATTGCTCAGCGG 3') primers were used.

1. Make a master mix of PCR reagent. For each sample to be amplified add: 5 µl 10X HotstarTaq buffer, 5 µl dNTP mix (2 mM each), 2.5 µl 5 µM T7 promoter primer, 2.5 µl 5 µM T7 terminator primer, 35 µl sterile H_20, and 0.25 µl (1.25 U) HoststarTaq polymerase.
2. Aliquot 50 µl of the master mix into each well of the PCR plate. Pick large colonies with sterile pipet tips and add to the appropriate wells.
3. Cycle under the following conditions: 95 °C 15 min, 94 °C 1 min, 55 °C 1 min, 72 °C 1 min (repeat steps 2-4 35 X), 72 °C 7 min.
4. Run 5 µl of each reaction on a 1% agarose gel to check for the presence of insert DNA of the proper size.

4. Notes

1. Any restriction enzyme can be used at this step, depending on the sequences that one is using for LIC. The desired product is blunt-ended DNA, so choose a blunt-cutting enzyme if possible. Another useful trick is to choose an enzyme with a recognition sequence that occurs a distance from the cleavage site: this way the recognition sequence does not need to fit in with the requirements of the flanking sequences.

2. If you are using the restriction enzyme BseR I or another enzyme that produces overhangs, do not alter the DNA concentration in this step because activity of the mung bean nuclease that is added directly to this reaction is concentration dependent.

3. It is very important to separate the undigested or nicked plasmid DNA from the linear DNA at this step, since failure to do so produces background colonies when the DNA is used for LIC. Although this step is time consuming, 10 µg of purified vector DNA is enough to perform 250 reactions.

4. When an N-terminal affinity tag is used, it is beneficial to add the stop codon to the insert sequence so that no additional amino acids are added to the C-terminal end of the protein. If the same PCR product is going to be cloned into many vectors (with affinity tags at N-and C-terminal ends), the insert sequence cannot contain the stop codon. In this case, the LIC sequence can be altered to add a STOP codon in a position that minimizes extra tag sequences.

5. The PCR reactions must be buffer exchanged to remove unincorporated dNTPs, even if the PCR products are very pure.

6. We have found that cloning efficiency is reduced when increased volumes of purified PCR product are used, possibly due to carry-over of contaminants from the clean-up reaction. Therefore, it is best to have the PCR product as concentrated as possible. The

exact amount of DNA in this step can vary somewhat to allow for standardized pipeting volumes without compromising cloning efficiency.

7. Our LIC scheme uses dCTP in the vector sample and dGTP in the insert sample to create the single-stranded overhangs. This can vary depending on the LIC sequence that is being used.

8. The exact amount of DNA in this step can vary somewhat to allow for standardized pipeting volumes without compromising cloning efficiency.

Acknowledgements

The author thanks Michael Murphy, Peter Beernink, and Paul Richardson for helpful suggestions and critical reading of the manuscript. This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Livermore National Laboratory (No. W-7405-Eng-48), Lawrence Berkeley National Laboratory (No. DE-AC03-76SF00098), and Los Alamos National Laboratory (No. W-7405-ENG-36).

References

1. Aslanidis, C. and deJong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. **18**, 6069-6074.

2. Haun, R.S., Servanti, I.M. and Moss, J. (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. Biotechniques **13**, 515-518.

3. Aslanidis, C., de Jong, P.J. and Schmitz, G. (1994) Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products. PCR Methods Appl. **4**(3), 172-177.

4. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F., and Donnelly, M. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expression and Purification **25**, 8-15.

5. Novagen (Madison, WI)

6. Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**, 113-130.

7. Rosenberg, A.H., Lade, B.N., Chui, D., Lin, S., Dunn, J.J., and Studier, F.W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene **56**, 125-135.

8. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Meth. Enzymol. **185**, 60-89.

Fig 1. Construction of the LIC vector containing the N-terminal hexahistidine affinity tag. The new vector pNHis, encodes a protein with a N-terminal extension of 6 histidine residues followed by 6 additional amino acids that encode a Factor Xa cleavage site. A stop codon was added to the gene sequence so that the 3' LIC sequence did not add 6 extra amino acids to the C-terminus of the protein sequence.

Fig 2. Ligation –independent cloning using the pNHis vector. The bold type denotes the LIC sequences of the vector, which contains the BseR I recognition sequence (underlined). The complementary LIC sequences of the insert DNA are shown in italics.